Projection with Bases

Note Title

LECTURE 09

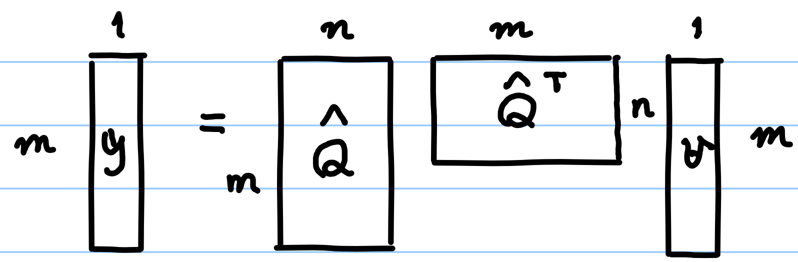
☆ Projection with an Orthonormal Basis

In general, for a projector $P \in \mathbb{R}^{m\times m}$
$\dim(\text{range}(P)) = n \le m$
So, let's define the matrix
$\hat{Q} := [q_1 \cdots q_n] \in \mathbb{R}^{m\times n}$
signifies the reduced form of Q
where $\{q_1, \cdots, q_n\}$ forms an O.N.B.
of range (P) as in the proof
of the previous theorem.
Then, $P = \hat{Q}\hat{Q}^T$ $\left(= Q \Lambda Q^T \text{ in the previous thm.} \right)$

Recall $\forall v \in \mathbb{R}^m$, $\exists r \in \mathbb{R}^m$ (residual)
s.t. $v = r + \sum_{i=1}^{n} (q_i q_i^T) v$

Hence, the mapping $v \longmapsto \sum_{i=1}^{n} (q_i q_i^T) v = y$

is an orthogonal projection onto range $(\hat{Q})$



as $n \to m$, $y \to v$ since $\hat{Q}\hat{Q}^T \to I$

Note: The complementary proj. to an orth. proj. $P = \hat{Q}\hat{Q}^T$ is also an orth. proj.

why? $I - P = I - \hat{Q}\hat{Q}^T$ is also symmetric! //

Also note the following special case:

The rank-one orth. proj. with a unit vector $q \in \mathbb{R}^m$

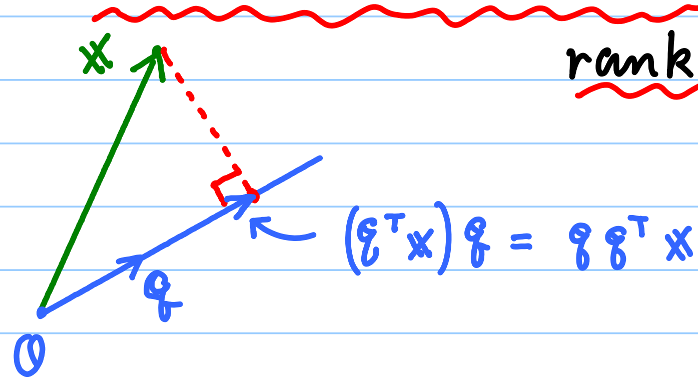
$$P_q := q q^T \in \mathbb{R}^{m \times m}$$

↳ a special rank 1 matrix

Its complementary proj. is

$$P_{\perp q} := I - P_q =: P_q^{\perp}$$

rank $m-1$ orth. proj.



For a general vector $a \in \mathbb{R}^m$ with $a \neq 0$, $\|a\| \neq 1$, the orth. proj. onto span$\{a\}$ becomes

$$P_a := \frac{a a^T}{a^T a} \rightarrow P_{\perp a} = I - \frac{a a^T}{a^T a}$$

why? Set $q = a / \|a\|$ then it's easy to show. //

☆ Projection with an Arbitrary Basis

Let $\{a_1, \cdots, a_n\} \subset \mathbb{R}^m$ be a set of linearly independent vectors.

Set $A = [a_1 \cdots a_n] \in \mathbb{R}^{m \times n}$ assume $m \geq n$.

Then what is the ortho. proj. onto $\text{range}(A) = \text{span}\{a_1, \cdots, a_n\}$?

Let $P_A \in \mathbb{R}^{m \times m}$ be such an ortho. proj.

Then $\text{range}(P_A) = \text{range}(A)$ and $\text{null}(P_A) = \text{null}(A^T)$.

Now take any $v \in \mathbb{R}^m$. Then $\exists v_1 \in \text{range}(A)$, $\exists v_2 \in \text{null}(A^T)$.

s.t. $v = v_1 + v_2$.

$\exists x \in \mathbb{R}^n$ s.t. $v_1 = Ax \perp v_2$

Hence $a_j \perp v_2 = v - v_1 = v - Ax$ for $1 \leq j \leq n$

$\Leftrightarrow \quad a_j^T (v - Ax) = 0 \quad 1 \leq j \leq n$

$\Leftrightarrow \quad A^T (v - Ax) = 0$

$\Leftrightarrow \quad A^T A x = A^T v \quad$ (normal eqn. !!)

HW03 Prob. 6 $\left\{\begin{array}{l} \text{Since } A \text{ is full rank (because } \{a_1, \cdots, a_n\} \text{ are lin. indep.)}, \\ (A^T A)^{-1} \text{ exists, i.e., } x = (A^T A)^{-1} A^T v \end{array}\right.$

Hence $P_A = A (A^T A)^{-1} A^T$

We recover the previous projectors by setting

$$\begin{cases} A = \hat{Q} \Rightarrow \hat{Q}(\underbrace{\hat{Q}^T\hat{Q}}_{=I.})^{-1}\hat{Q}^T = \hat{Q}\hat{Q}^T \\ A = \underbrace{[a]}_{m\times 1 \text{ matrix}} \Rightarrow a(\underbrace{a^Ta}_{\text{a scalar}})^{-1}a^T = \dfrac{aa^T}{a^Ta} \end{cases}$$

Ex. $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$

(1) Compute the orthogonal projector P_A onto range(A)

(2) Let $v = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Then compute its projection onto range(A) using P_A obtained in (1).

<u>Sol.</u> (1) $P_A = A(A^TA)^{-1}A^T$

If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then $\underbrace{A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}}$

$A^TA = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$

$(A^TA)^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}$

So, $P_A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}$ //

(2) $P_A v = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$ //

QR Decomposition Factorization

★ Reduced QR Factorization

Let $A = [a_1 \cdots a_n] \in \mathbb{R}^{m \times n}$

Denote $\text{span}\{a_1, \cdots, a_k\}$ by $\langle a_1, \cdots, a_k \rangle$ for simplicity.

Then $\langle a_1 \rangle \subset \langle a_1, a_2 \rangle \subset \langle a_1, a_2, a_3 \rangle \subset \cdots \subset \langle a_1, \cdots, a_n \rangle$

successive column spaces.

QR factorization (or decomposition) := a method to successively construct a sequence of orthonormal vectors $q_1, q_2, \cdots$, s.t.

$$\langle q_1, \cdots, q_j \rangle = \langle a_1, \cdots, a_j \rangle$$

$$j = 1, 2, \cdots, n$$

If this is the case, then each a_j can be expanded by $\{q_1, \cdots, q_j\}$

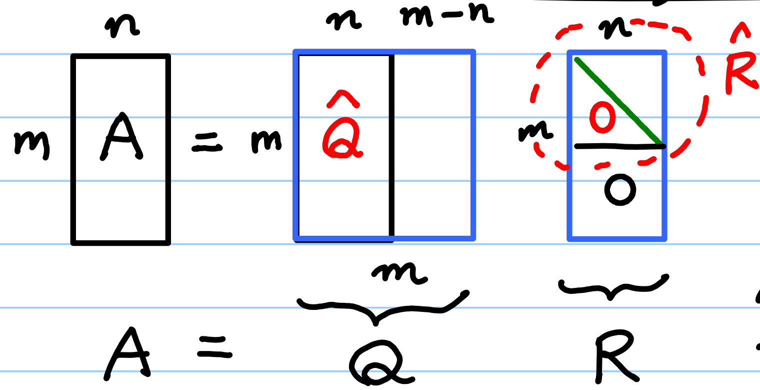
Say,
$$\begin{cases} a_1 = r_{11} q_1 \\ a_2 = r_{12} q_1 + r_{22} q_2 \\ \vdots \\ a_n = r_{1n} q_1 + r_{2n} q_2 + \cdots + r_{nn} q_n \end{cases}$$

$$\Rightarrow A = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & 0 & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}$$

$$= \hat{Q} \hat{R}$$

Reduced QR factorization of A — upper triangular

★ Full QR Factorization



Append $m-n$ O.N. vectors to col's of $\hat{Q} \to Q$

Append $m-n$ 0's to rows of $\hat{R} \to R$

Note: $q_j \perp \text{range}(A)$ for $j > n$.

i.e., $\langle q_1, \cdots, q_n \rangle = \text{range}(A)$

$\langle q_{n+1}, \cdots, q_m \rangle = \text{range}(A)^\perp$

$= \text{null}(A^T)$

★ The Classical Gram-Schmidt Orthogonalization as QR

You must be familiar with the classical GS procedure.

Given $\{a_1, \cdots, a_n\}$, $a_j \in \mathbb{R}^m$, $1 \le j \le n$, construct an orthonormal set $\{q_1, \cdots, q_n\}$, $q_j \in \mathbb{R}^m$, $1 \le j \le n$ as follows:

$$q_1 = \frac{a_1}{r_{11}}, \quad r_{11} = \|a_1\|$$

$$q_2 = \frac{a_2 - r_{12} q_1}{r_{22}}, \quad r_{12} = q_1^T a_2, \quad r_{22} = \|a_2 - r_{12} q_1\|$$

$$q_3 = \frac{a_3 - r_{13} q_1 - r_{23} q_2}{r_{33}}, \quad r_{i3} = q_i^T a_3,\ i=1,2. \quad r_{33} = \| a_3 - r_{13} q_1 - r_{23} q_2 \|$$

$$\vdots \qquad\qquad\qquad \vdots$$

$$q_n = \frac{a_n - \sum_{i=1}^{n-1} r_{in} q_i}{r_{nn}}, \quad r_{nn} = \| a_n - \sum_{i=1}^{n-1} r_{in} q_i \|$$

So, in general,

$$\begin{cases} r_{ij} = q_i^T a_j & i \neq j \\ r_{jj} = \| a_j - \sum_{i=1}^{j-1} r_{ij} q_i \| \end{cases}$$

Algorithm (The classical Gram-Schmidt) (CGS)

for $j = 1 : n$
- $v_j = a_j$
- for $i = 1 : j-1$
  - $r_{ij} = q_i^T a_j$
  - $v_j = v_j - r_{ij} q_i$ ← error accumulates here
- $r_{jj} = \| v_j \|$
- $q_j = v_j / r_{jj}$

Note the order of r_{ij} computation

j:	1	2	3	....	n
	r_{11}	r_{12}	r_{13}		r_{1n}
		r_{22}	r_{23}		r_{2n}
			r_{33}		r_{3n}
					$\vdots$
					r_{nn}

Unfortunately, this version is numerically unstable.

Ex. $A = \begin{bmatrix} 1 & 1 & 1 \\ \varepsilon & 0 & 0 \\ 0 & \varepsilon & 0 \\ 0 & 0 & \varepsilon \end{bmatrix}$. ε : small s.t. ε^2 can be ignored.

Apply CGS.

$$\begin{cases} r_{11} = \|a_1\| = \sqrt{1^2 + \varepsilon^2 + 0^2 + 0^2} \approx 1. \\ q_1 = \dfrac{a_1}{r_{11}} = a_1. \end{cases}$$

$$\begin{cases} r_{12} = q_1^T a_2 = [1\ \varepsilon\ 0\ 0] \begin{bmatrix} 1 \\ 0 \\ \varepsilon \\ 0 \end{bmatrix} = 1. \\ q_2 = \dfrac{a_2 - r_{12} q_1}{r_{22}} = \dfrac{1}{r_{22}} \left(\begin{bmatrix} 1 \\ 0 \\ \varepsilon \\ 0 \end{bmatrix} - 1 \cdot \begin{bmatrix} 1 \\ \varepsilon \\ 0 \\ 0 \end{bmatrix} \right) = \dfrac{1}{r_{22}} \begin{bmatrix} 0 \\ -\varepsilon \\ \varepsilon \\ 0 \end{bmatrix} \\ r_{22} = \|a_2 - r_{12} q_1\| = \varepsilon \sqrt{0 + (-1)^2 + 1^2 + 0^2} = \sqrt{2}\,\varepsilon \\ q_2 = \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \end{cases}$$

$$\begin{cases} r_{13} = q_1^T a_3 = 1, \quad r_{23} = q_2^T a_3 = 0. \\ q_3 = \dfrac{a_3 - r_{13} q_1 - r_{23} q_2}{r_{33}} = \dfrac{1}{r_{33}} \left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ \varepsilon \end{bmatrix} - \begin{bmatrix} 1 \\ \varepsilon \\ 0 \\ 0 \end{bmatrix} \right) \\ \quad = \dfrac{1}{r_{33}} \begin{bmatrix} 0 \\ -\varepsilon \\ 0 \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \text{ with } r_{33} = \sqrt{2}\,\varepsilon. \end{cases}$$

Hence

$$\hat{Q} = \begin{bmatrix} 1 & 0 & 0 \\ \varepsilon & -1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 0 \\ 0 & 0 & 1/\sqrt{2} \end{bmatrix}, \quad \hat{R} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & \sqrt{2}\varepsilon & 0 \\ 0 & 0 & \sqrt{2}\varepsilon \end{bmatrix}$$

Let's check these results.

$$\hat{Q}\hat{R} = \begin{bmatrix} 1 & 1 & 1 \\ \varepsilon & 0 & 0 \\ 0 & \varepsilon & 0 \\ 0 & 0 & \varepsilon \end{bmatrix} = A$$

So, looks OK.

But, how about the orthogonality of $\hat{Q}$?

$$\hat{Q}^T\hat{Q} = \begin{bmatrix} 1 & \varepsilon & 0 & 0 \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ \varepsilon & -1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 0 \\ 0 & 0 & 1/\sqrt{2} \end{bmatrix}$$

$$= \begin{bmatrix} 1+\varepsilon^2 & -\varepsilon/\sqrt{2} & -\varepsilon/\sqrt{2} \\ -\varepsilon/\sqrt{2} & 1 & 1/2 \\ -\varepsilon/\sqrt{2} & 1/2 & 1 \end{bmatrix}$$

$$\approx \begin{bmatrix} 1 & -\varepsilon/\sqrt{2} & -\varepsilon/\sqrt{2} \\ -\varepsilon/\sqrt{2} & 1 & 1/2 \\ -\varepsilon/\sqrt{2} & 1/2 & 1 \end{bmatrix}$$

$$\neq I_{3\times 3}$$

This is called "loss of orthogonality".